

WHAT IS CLAIMED IS:

- 1 1. A method comprising:
 - 2 accepting text spellings of training words in a plurality of sets of training words, each
 - 3 set corresponding to a different one of a plurality of languages;
 - 4 for each of the sets of training words in the plurality, receiving pronunciations for the
 - 5 training words in the set, the pronunciations being characteristic of native speakers of the
 - 6 language of the set, the pronunciations also being in terms of subword units at least some of
 - 7 which are common to two or more of the languages; and
 - 8 training a single pronunciation estimator using data comprising the text spellings and
 - 9 the pronunciations of the training words.
- 1 2. The method of claim 1 further comprising:
 - 2 accepting a plurality of sets of utterances, each set corresponding to a different one of
 - 3 the plurality of languages, the utterances in each set being spoken by the native speakers of
 - 4 the language of each set; and
 - 5 training a set of acoustic models for the subword units using the accepted sets of
 - 6 utterances and pronunciations estimated by the single pronunciation estimator from text
 - 7 representations of the training utterances.
- 1 3. The method of claim 1, wherein a first training word in a first set in the plurality
 - 2 corresponds to a first language and a second training word in a second set corresponds to a
 - 3 second language, the first and second training words having identical text spellings, the
 - 4 received pronunciations for the first and second training words being different.
- 1 4. The method of claim 3, wherein utterances of the first and the second training words are
 - 2 used to train a common subset of subword units.
- 1 5. The method of claim 1, wherein the single pronunciation estimator uses a decision tree to
 - 2 map letters of the text spellings to pronunciation subword units.
- 1 6. The method of claim 1, where training the single pronunciation estimator further
 - 2 comprises:

3 forming, from sequences of letters of each training word's textual spelling and the
4 corresponding grouping of subword units of the pronunciation, a letter to subword mapping
5 for each training word; and
6 training the single pronunciation estimator using the letter-to-subword mappings.

1 7. The method of claim 6, wherein training the single pronunciation estimator and training
2 the acoustic models is executed by a nonportable programmable device.

1 8. The method of claim 1 further comprising:

2 generating, for each word in a list of words to be recognized, an acoustic word model,
3 the generating comprising generating a grouping of subword units representing a
4 pronunciation of the word to be recognized using the single pronunciation estimator.

1 9. The method of claim 8, wherein the grouping of subword units is a linear sequence of
2 subword units.

1 10. The method of claim 9, wherein the grouping of the acoustic subword models is a linear
2 sequence of acoustic subword models.

1 11. The method of claim 8, wherein the subword units are phonemes.

1 12. The method of claim 8, wherein the grouping of subwords is a network, and the network
2 represents two pronunciations of a word, the two pronunciations being representative of
3 utterances of native speakers of two languages.

1 13. The method of claim 8 further comprising:

2 processing an utterance; and

3 scoring matches between the processed utterance and the acoustic word models.

1 14. The method of claim 13, wherein generating the acoustic word model, processing the
2 utterance, and scoring matches is executed by a portable programmable device.

1 15. The method of claim 14, wherein the portable programmable device is a cellphone.

1 16. The method of claim 13, wherein the utterance is spoken by a native speaker of one of the
2 plurality of languages.

1 17. The method of claim 14, wherein the utterance is spoken by a native speaker of a
2 language other than the plurality of languages, the language having similar sounds and
3 similar letter to sounds rules as a language from the plurality of languages.

1 18. A method for recognizing words spoken by native speakers of multiple languages, the
2 method comprising:

3 generating a set of estimated pronunciations, using a single pronunciation estimator,
4 from text spellings of a set of acoustic training words, each pronunciation comprising a
5 grouping of subword units, the set of acoustic training words comprising at least a first word
6 and a second word, the first and second words having identical text spelling, the first word
7 having a pronunciation based on utterances of native speakers of a first language, the second
8 word having a pronunciation based on utterances of native speakers of a second language;

9 mapping sequences of sound associated with utterances of each of the acoustic
10 training words against the estimated pronunciation associated with each of the acoustic
11 training words; and

12 using the mapping of sequences of sound to estimated pronunciations to generate
13 acoustic subword models for the subword units in the grouping of subwords, the acoustic
14 subword model comprising a sound model and a subword unit.

1 19. A method for multilingual speech recognition comprising:

2 accepting a recognition vocabulary that includes words from multiple languages;

3 determining a pronunciation of each of the words in the recognition vocabulary using
4 a pronunciation estimator that is common to the multiple languages; and

5 configuring a speech recognizer using the determined pronunciations of the words in
6 the recognition vocabulary.

1 20. The method of claim 19 further comprising:

2 accepting a training vocabulary that comprises words from multiple languages;

3 determining a pronunciation of each of the words in the training vocabulary using the
4 pronunciation estimator that is common to the multiple languages;

5 configuring the speech recognizer using parameters estimated using the determined
6 pronunciations of the words in the training vocabulary; and

7 recognizing utterances using the configured speech recognizer.

1 21. A computer program product, tangibly embodied in an information carrier, the computer
2 program product being operable to cause data processing apparatus to:

3 accept text spellings of training words in a plurality of sets of training words, each set
4 corresponding to a different one of a plurality of languages;

5 for each of the sets of training words in the plurality, receive pronunciations for the
6 training words in the set, the pronunciations being characteristic of native speakers of the
7 language of the set, the pronunciations also being in terms of subword units at least some of
8 which are common to two or more of the languages; and

9 train a single pronunciation estimator using data comprising the text spellings and the
10 pronunciations of the training words.

1 22. The computer program product of claim 21, the computer program product being further
2 operable to cause the data processing apparatus to:

3 accept a plurality of sets of utterances, each set corresponding to a different one of the
4 plurality of languages, the utterances in each set being spoken by the native speakers of the
5 language of each set; and

6 train a set of acoustic models for the subword units using the accepted sets of
7 utterances and pronunciations estimated by the single pronunciation estimator from text
8 representations of the training utterances.

1 23. The computer program product of claim 22, wherein a first training word in a first set in
2 the plurality corresponds to a first language and a second training word in a second set
3 corresponds to a second language, the first and second training words having identical text
4 spellings, the received pronunciations for the first and second training words being different.

1 24. The computer program product of claim 23, wherein utterances of the first and the second
2 training words are used to train a common subset of subword units.

1 25. The computer program product of claim 21, wherein the single pronunciation estimator
2 uses a decision tree to map letters of the text spellings to pronunciation subword units.

1 26. The computer program product of claim 21, wherein training the single pronunciation
2 estimator further comprises:

3 form, from sequences of letters of each training word's textual spelling and the
4 corresponding grouping of subword units of the pronunciation, a letter to subword mapping
5 for each training word; and
6 train the single pronunciation estimator using the letter-to-subword mappings.

1 27. The computer program product of claim 22, wherein training the single pronunciation
2 estimator and training the acoustic models is executed by a nonportable programmable
3 device.

1 28. The computer program product of claim 22, the computer program product being further
2 operable to cause the data processing apparatus to:
3 generate, for each word in a list of words to be recognized, an acoustic word model,
4 the generating comprising generating a grouping of subword units representing a
5 pronunciation of the word to be recognized using the single pronunciation estimator.

1 29. The computer program product of claim 28 wherein the grouping of subword units is a
2 linear sequence of subword units.

1 30. The computer program product of claim 29, wherein the grouping of the acoustic
2 subword models is a linear sequence of acoustic subword models.

1 31. The computer program product of claim 28, wherein the subword units are phonemes.

1 32. The computer program product of claim 28, wherein the grouping of subwords is a
2 network, and the network represents two pronunciations of a word, the two pronunciations
3 being representative of utterances of native speakers of two languages.

1 33. The computer program product of claim 28, the computer program product being further
2 operable to cause the data processing apparatus to:

3 process an utterance; and

4 score matches between the processed utterance and the acoustic word models.

1 34. The computer program product of claim 33, wherein generating the acoustic word model,
2 processing the utterance, and scoring matches is executed by a portable programmable
3 device.

1 35. The computer program product of claim 34, wherein the portable programmable device is
2 a cellphone.

1 36. The computer program product of claim 33, wherein the utterance is spoken by a native
2 speaker of one of the plurality of languages.

1 37. The computer program product of claim 35, wherein the utterance is spoken by a native
2 speaker of a language other than the plurality of languages, the language having similar
3 sounds and similar letter to sounds rules as a language from the plurality of languages.

1 38. A computer program product for recognizing words spoken by native speakers of
2 multiple languages, the computer program product being operable to cause data processing
3 apparatus to:

4 generate a set of estimated pronunciations, using a single pronunciation estimator,
5 from text spellings of a set of acoustic training words, each pronunciation comprising a
6 grouping of subword units, the set of acoustic training words comprising at least a first word
7 and a second word, the first and second words having identical text spelling, the first word
8 having a pronunciation based on utterances of native speakers of a first language, the second
9 word having a pronunciation based on utterances of native speakers of a second language;

10 map sequences of sound associated with utterances of each of the acoustic training
11 words against the estimated pronunciation associated with each of the acoustic training
12 words; and

13 use the mapping of sequences of sound to estimated pronunciations to generate
14 acoustic subword models for the subword units in the grouping of subwords, the acoustic
15 subword model comprising a sound model and a subword unit.

1 39. A computer program product for multilingual speech recognition, the computer program
2 product being operable to cause data processing apparatus to:

3 accept a recognition vocabulary that includes words from multiple languages;

4 determine a pronunciation of each of the words in the recognition vocabulary using a
5 pronunciation estimator that is common to the multiple languages; and

6 configure a speech recognizer using the determined pronunciations of the words in
7 the recognition vocabulary.

1 40. The computer program product of claim 40, the computer program product being further
2 operable to cause data processing apparatus to:

3 accept a training vocabulary that comprises words from multiple languages;
4 determine a pronunciation of each of the words in the training vocabulary using the
5 pronunciation estimator that is common to the multiple languages;
6 configure the speech recognizer using parameters estimated using the determined
7 pronunciations of the words in the training vocabulary; and
8 recognize utterances using the configured speech recognizer.

1 41. An apparatus comprising:

2 means for accepting text spellings of training words in a plurality of sets of training
3 words, each set corresponding to a different one of a plurality of languages;

4 means for receiving, for each of the sets of training words in the plurality,
5 pronunciations for the training words in the set, the pronunciations being characteristic of
6 native speakers of the language of the set, the pronunciations also being in terms of subword
7 units at least some of which are common to two or more of the languages; and

8 means for training a single pronunciation estimator using data comprising the text
9 spellings and the pronunciations of the training words.

1 42. The apparatus of claim 41 further comprising:

2 means for accepting a plurality of sets of utterances, each set corresponding to a
3 different one of the plurality of languages, the utterances in each set being spoken by the
4 native speakers of the language of each set; and

5 means for training a set of acoustic models for the subword units using the accepted
6 sets of utterances and pronunciations estimated by the single pronunciation estimator from
7 text representations of the training utterances.

1 43. The apparatus of claim 42 further comprising:

2 a means for generating, for each word in a list of words to be recognized, an acoustic
3 word model, the generating comprising generating a grouping of subword units representing
4 a pronunciation of the word to be recognized using the single pronunciation estimator.

1 44. The apparatus of claim 43 further comprising:

2 means for processing an utterance; and
3 means for scoring matches between the processed utterance and the acoustic word
4 models.

1 45. An apparatus for recognizing words spoken by native speakers of multiple languages, the
2 apparatus comprising:

3 a means for generating a set of estimated pronunciations, using a single pronunciation
4 estimator, from text spellings of a set of acoustic training words, each pronunciation
5 comprising a grouping of subword units, the set of acoustic training words comprising at
6 least a first word and a second word, the first and second words having identical text spelling,
7 the first word having a pronunciation based on utterances of native speakers of a first
8 language, the second word having a pronunciation based on utterances of native speakers of a
9 second language;

10 means for mapping sequences of sound associated with utterances of each of the
11 acoustic training words against the estimated pronunciation associated with each of the
12 acoustic training words; and

13 means for using the mapping of sequences of sound to estimated pronunciations to
14 generate acoustic subword models for the subword units in the grouping of subwords, the
15 acoustic subword model comprising a sound model and a subword unit.

1 46. An apparatus for multilingual speech recognition, the apparatus comprising:

2 means for accepting a recognition vocabulary that includes words from multiple
3 languages;

4 means for determining a pronunciation of each of the words in the recognition
5 vocabulary using a pronunciation estimator that is common to the multiple languages; and

6 means for configuring a speech recognizer using the determined pronunciations of the
7 words in the recognition vocabulary.